Active Object Detection and Pose Estimation in General Belief Space

Kanzhi Wu, Teng Zhang, Gamini Dissanayake, Shoudong Huang and Ravindra Ranasinghe
Centre for Autonomous Systems, University of Technology Sydney
kanzhi.wu,teng.zhang,gamini.dissanayake,shoudong.huang,ravindra.ranasinghe@uts.edu.au

I. INTRODUCTION AND RELATED WORK

3D object detection and pose estimation is an indispensable task for service robots, however, due to the challenges in everyday environments where occlusion are observed frequently, it has been recognised that object detection and localisation cannot be achieved using observation from a single viewpoint. Active object detection, recognising the targeting objects via active maneuver, is a solution towards improving the recognition probability and minimising the estimation uncertainty through fully utilising the mobility of the robotic platform. Active perception, proposed by Bajcsy [2], is defined as the problem of intelligent control strategies applied to the data acquisition process which will depend on the current state of data interpretation including recognition. Active object detection and pose estimation, as a subset of active perception, has been investigated intensively in recent years. Atanasov and al. [1] introduced Viewpoint-Pose tree (VP-Tree) which provides not only object category but also its relative pose and the active hypothesis testing problem is addressed by a point-based POMDP algorithm. Potthast and et al. [4] combined next-best-view planning and feature selection into one online active object detection system. Based on simple features and detectors, their algorithm enables the quadrotor to detect the active object detection and pose estimation, as a subset of active perception, acquisition process which will depend on the current state of data interpretation including recognition. Active object detection, recognising the targeting objects via active maneuver, is a solution towards improving the recognition probability and minimising the estimation uncertainty through fully utilising the mobility of the robotic platform. Active perception, proposed by Bajcsy [2], is defined as the problem of intelligent control strategies applied to the data acquisition process which will depend on the current state of data interpretation including recognition. Active object detection and pose estimation, as a subset of active perception, has been investigated intensively in recent years. Atanasov and al. [1] introduced Viewpoint-Pose tree (VP-Tree) which provides not only object category but also its relative pose and the active hypothesis testing problem is addressed by a point-based POMDP algorithm. Potthast and et al. [4] combined next-best-view planning and feature selection into one online active object detection system. Based on simple features and detectors, their algorithm enables the quadrotor to detect the targeting object more robustly.

In this paper, discretisation on state and control spaces are not required and similar as Indelman’s work [3], based on the probabilistic representation of the general belief, the next-best-view are further decided via optimising the objective function parameterised by the general belief, thus enabling the planning in the continuous domain. We follow the maximum likelihood observation assumption and construct the objective function which consolidates uncertainties of the robot and the poses of detected objects, object detection confidences and control consumption. The simulation results demonstrate the effectiveness of the proposed approach. The conclusion and future work are discussed in section. [IV]

II. NEXT-BEST-VIEW DECISION IN GENERAL BELIEF SPACE

A. Preliminaries and problem formulation

The parameters which will be used in following sections are listed in table. [I] and the noises appeared in the table are assumed to be Gaussian. In table. [I] the motion model and observation model can be further simplified. For motion model $G$, without acquiring observation, object states $[X^0_t, X^2_t, ..., X^m_t]$ in $X_k$ will not be updated and only the robot pose $X^r$ will be computed according to:

$$X^r_k = g \left(X^r_{k-1}, u_{k-1} \right) + \eta_r$$

(1)

Also, the observation $Z_k$ is only decided by the current pose of the robot $X^r_k$ and estimated poses of the objects $[X^0_t, ..., X^m_t]$ thus $H$ can be simplified as:

$$Z_k = H \left(X^r_k, X^r_k, X^r_k \right) + \xi$$

(2)

Considering each feature individually, we have:

$$z_{k,j} = h \left(X^o, X^r_k, p^f_j \right) + \xi_j$$

(3)

where $p^f_j$ is the coordinate of feature $j$ in object frame. Similar as Indelman’s work, the generalised belief in $l$-th predication step is denoted as:

$$b \left(X_{k+l} \right) \equiv p \left(X_{k+l} | Z_k, u_{k-1}, Z_{k+1:k+l}, u_{k:k+l-1} \right)$$

(4)

and this belief is assumed to be Gaussian as:

$$b \left(X_{k+l} \right) \sim \mathcal{N} \left(X^\Delta_{k+l}, I_{k+l} \right)$$

(5)

where $X^\Delta_{k+l}$ is the maximum a posterior (MAP) estimation and $I_{k+l}$ is the corresponding information matrix. In order to find the optimal control inputs $u_{k:k+l-1}$ in next $L-1$ steps, we need to minimise the objective function $J_k$ parameterised by $u_{k:k+l-1}$ and the detailed formulation of $J_k$ will be illustrated in the following section.

$$u^*_{k:k+l-1} = \arg \min_{u_{k:k+l-1}} J_k (u_{k:k+l-1})$$

(6)

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^o$</td>
<td>object pose</td>
</tr>
<tr>
<td>$X^r$</td>
<td>robot pose</td>
</tr>
<tr>
<td>$X_k$</td>
<td>$X_k = [X^0_k, ..., X^m_k]$ which is the state of time step $k$, $i$ is the number of detected objects up to time step $k$</td>
</tr>
<tr>
<td>$z_{k,j}$</td>
<td>observation of feature $j$ at time step $k$</td>
</tr>
<tr>
<td>$Z_k$</td>
<td>$Z_k = [z_{k,1}, z_{k,2}, ..., z_{k,N_k}]$ where $N_k$ is the number of observed features at time step $k$</td>
</tr>
<tr>
<td>$Z_{k-1}$</td>
<td>control input at time step $k-1$</td>
</tr>
<tr>
<td>$u_{k-1}$</td>
<td>$u_{k-1} = [u_1, u_2, ..., u_{k-1}]$</td>
</tr>
<tr>
<td>$G$</td>
<td>$G \left(X_{k-1}, u_{k-1} \right)$</td>
</tr>
<tr>
<td>$H$</td>
<td>$H \left(X^r, X^r, X^r \right) + \xi$ is the observation model</td>
</tr>
</tbody>
</table>

TABLE I: Summary of Notations
B. Objective function formulation

In order to compute the optimal control $u_{k:k+L-1}$ via optimising the objective function $J_k$, we can divide the problem into 2 steps: 1) inference $b(X_{k:L})$ given control inputs $u_{k:k+L-1}$ and 2) compute the $u_{k:k+L-1}^{(i)}$ which minimises the objective function. The 2 steps are executed iteratively until the stop criterion are met. The second step can be realised by gradient based optimisation methods and the control inputs will be updated as:

$$u_{k:k+L-1}^{(i+1)} = u_{k:k+L-1}^{(i)} - \lambda \nabla J_k$$

where $\nabla J_k$ is the gradient of objective function at guess $u_{k:k+L-1}^{(i)}$. The main difficulty lies in predicting the general belief $b(X_{k:t})$ without acquiring the observations. Following by the derivations in [3] but with slightly difference (assuming maximum likelihood observation), using Expectation-Maximisation and formulating the objective function as a quadratic form, we approximate the $X_{k+1}$ and $I_{k+1}^{(i)}(u_{k:k+L-1})$:

$$X_{k+1} = X_{k+1}$$
$$I_{k+1} = A_{k+1}^T A_{k+1}$$

where $X_{k+1}$ is the nominal state and $G$ and $H$ are stacked by the Jacobians for motion model and observation model.

The objective function shown below consists 3 terms (costs):

$$J_k(u_{k:k+L-1}) = \text{Tr} \left( M_{\Sigma}^{-1} \sum_{i=1}^{n_{obj}} \psi(\mathbf{Y}_{n_i | X_{k:L}}) \right) + \sum_{l=0}^{L-1} \Phi(u_l)$$

1) Uncertainties of the robot pose and the poses of detected objects. In eq. [9] the uncertainty is obtained by calculating the trace(weighted trace) of the covariance matrix as $\sum_{l=0}^{L-1} \text{Tr} (M_{\Sigma}^{-1} I_{k+l} M_{\Sigma}^T)$;

2) Object detection confidences parameterised by observed features. Based on the MAP estimate of $X_{k:L}$, we construct a observation score matrix $\mathbf{Y}_{n_i | X_{k:L}}$ which describes the feature matching score under a specific relative pose where $n_i$ is the number of features in object $i$. We compute the observation score $v = \psi(\mathbf{Y})$ which describes the probability of missing the targeting objects after a sequence of observations;

3) Control constraints where larger control inputs will be penalised. In eq. [9] the cost function is denoted as $\Phi(u_l)$. In section. III we assign larger cost for larger angular velocity control input;

III. EXPERIMENTS AND RESULTS

We present simulation results, the planned trajectories for active object detection, using the proposed approach in fig. 1. The first row in fig. 1 shows the results for single object and the second row shows the results for multiple objects. The green trajectories are generated from the initial control input, the red ones are computed from the optimised control input and executed. During the simulation experiments, we consider the further 5 steps in the planning phase. The robot only executes one step after planning for 5 steps ahead, therefore, for each term in the objective function, larger weight is associated with nearer steps which indicates that the robot are more interested in short-term gain rather than the long term gain. The state, the robot pose and object pose, is updated using EKF after acquiring the updated observation and the uncertainty of the poses are reduced after observing the object for multiple times. Fig. 2 demonstrates the uncertainty of the estimated object corresponding to the two scenarios in fig. 1.

Fig. 1: Active object detection trajectories

Fig. 2: Object uncertainty before and after active planning

IV. CONCLUSION AND FUTURE WORK

This paper presents a probabilistic framework for active object detection and pose estimation which is able to provide the optimal control in continuous space. The simulation results demonstrates the effectiveness of the proposed approach. The remaining difficulty is eliminating the maximum likelihood observation assumption and considers the unexpected occlusion or noise during the prediction. The experiments on actual robotic platform will be conducted in future work.

REFERENCES


